STATISTICAL COMMISSION and Working Paper No. 23
ECONOMIC COMMISSION FOR EUROPE English only

CONFERENCE OF EUROPEAN STATISTICIANS

Joint UNECE/EUROSTAT Work Session on Methodological Issues Involving the Integration of Statistics and Geography
(Tallinn, Estonia, 25-28 September 2001)

Topic (iii): Spatial analysis

THE DEVELOPMENT OF DELINEATION METHODS OF URBAN AREAS FOR THE CENSUS 2000

Submitted by Statistics Finland[1]

Contributed paper

ABSTRACT

In Finland there are 5.2 million inhabitants and the land area is 305 000 km2. This represents about 17 people per km2 of land area. There are only a few concentrations of inhabitants that can be characterised as an 'urban' area in a European sense.

In sparsely populated Finland the concept of locality acquired its everyday meaning of population centre long before the term was given a precise definition. Locality was used as a statistical area in 1950, in the first general census covering the entire country. But the history of locality statistics has its actual beginning in the 1960 census, when the common Nordic definition of locality, adopted the same year, was applied for the first time. Since then, localities have been delineated in connection with all censuses. Through to 1985, the boundaries were drawn on basic maps. In 1990 and 1995 the delimitation method was computerized and based on the coordinate data of buildings.

For the 2000 census, the Geographic Information Team of Statistics Finland developed and implemented a method for delineation of localities in GIS. The method is based on the register data of buildings and follows the Nordic definition of localities. This paper is based on the studies and experiences of the Team and focuses on problems and questions noted by the Team during the process of developing.

The first part of this paper describes some requirements and bases for the delineation method, such as the definition of localities and the basic data to be used. The development of the delineation process is discussed in the second part of the paper. The third part contains some evaluation of the work process and the final delineation.

I. REQUIREMENTS AND BASES FOR THE DELINEATION METHOD

I.1 Definition of localities from the year 1960

1. Localities are clusters of buildings housing at least 200 residents, where the distance between buildings normally does not exceed 200 meters. The distance can be in excess of 200 meters if the cluster of buildings lies within an area belonging to the sphere of influence of a major urban settlement. On the other hand, the maximum distance between buildings must be less than 200 meters if the nature of the

[1] Prepared by Liisa Kanerva.

settlement so requires. This is the case, for example, when the border between localities and sparsely populated area is unclear, i.e. when the locality is not significantly more densely populated than the surrounding area.

2. The delineation of localities also takes into account buildings other than residential buildings, i.e. buildings used exclusively for work places. However, buildings used for agricultural production are not included in a locality if they are situated away from the main buildings of the farm. Hospitals and other institutions situated outside localities are included only if the number of their resident staff, including family members, is at least 200. Hospital patients are not taken into account.

3. A distance in excess of 200 meters between buildings does not break the continuity of an urban settlement if the area between buildings is used for public purposes (e.g. road, car park, park, athletic field, railway lines and harbour areas). On the other hand, an area with free time residences not permanently occupied is not included in a locality.

4. The delimitation of localities is not restricted by any administrative area divisions.

I.2 Input data

5. The basic data included all the buildings except free-time residences with their location, the number of resident population and the use of the building. The building stock of the year 2000 is about 1,300,000 buildings in all, 97.8 per cent of which are covered by coordinates. The location is geocoded in the central point of the buildings to an accuracy of one meter . The shape or the corner points of the buildings were not available. Residents are linked to the buildings by a domicile code, which specifies the exact location of the person´s residence.

II. DEVELOPING THE METHOD

6. In 1998 Statistics Finland carried out a project to produce small area statistics on rural villages of Finland. The project delineated all building blocks from one building to large urban areas for that purpose. Because there is no common or harmonised definition of rural area at international or national levels in Finland, the spatial parameters were settled by the local authority experts and regional planning associations. The method adapted could be referred to as the grid and buffer method. It has been developed with functions and commands inside the ArcInfo software and it is quite simple and very efficient in terms of costs and resources.

II.1 Grid and buffer method

7. Briefly described, the method consists of three tasks. First, clustering by converting points to raster surface (grid) where the value of a pixel (grid cell, square) is information if, inside the pixel area, there are no buildings, only residential buildings. The second task is to convert pixels to polygons and finally to merge the polygons by buffering. The buffer zone guarantees that all buildings, even those large in size and situated on the side of the grid cell, are inside the boundary line.

II.2 Functional solutions

8. As the first task, two grids were converted: one with all buildings and the other with only non-residential buildings. The width of the buffer zone destined for grid cells including only residential buildings (e.g. 50) was assigned as the value of the built-up grid cell in the first grid. The width of the buffer zone used for grids with large buildings (e.g. 25) was assigned as the value of the built-up grid cell in the second grid. A grid cell with no data was given the value zero in both grids. When those two grids are summed up, the cell value of the new grid is the width of the buffer zone with which each cell is meant to be buffered (e.g. 50 or 75).

9. In the second task the grid was converted into polygons. The ArcInfo function GRIDPOLY creates arcs from cell borders in the grid. Contiguous grid cells with the same value are grouped together to form polygons. The polygons have label points and a polygon attribute table (PAT). The PAT contains an item used to store the value of corresponding grid cells (Figure1).

10. The third task consists of buffering the polygons with the distance value stored in PAT. Except for the formation of an exclusion area, buffering merges the adjacent polygons together if they are closer than two times the value of distance from one another (Figure1).

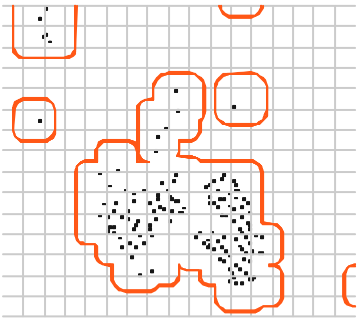

Figure 1. Buildings, grid and buffered grid polygons

11. In the final task, the residential people inside boundaries were identified and their number was calculated. Areas with 200 people were picked out (Figure 2). To find those building blocs in close contact with a locality, a buffer zone was formed and the polygons inside or touching the buffer line were picked out (Figure 2). Then those polygons were merged to the main area by buffering (Figure 3). To obtain the final form of the locality, the buffer zone was resized back to its original dimension (Figure 4).

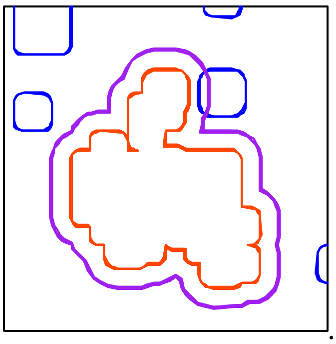

Figure 2. Selecting building blocks inside the area of influence of a locality

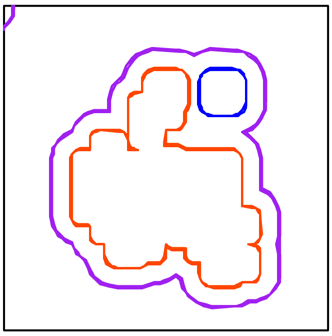

Figure 3. Merging polygons together by buffering

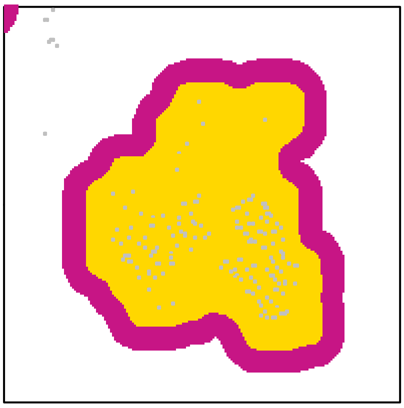

Figure 4. Resizing buffer zone back to original

II.3 Choosing spatial parameters to fulfil the definition of locality

12. By increasing and decreasing the resolution of a grid and the width of the buffer zone, one can delineate more accurately or generously. In our case, the distance between buildings should not exceed 200 metres. That means that there were many possibilities for gridding and buffering. The 50 metre resolution with a 100 metre buffer zone, the 75 metre resolution with 50- and 75 metre buffer zones and the 100 metre resolution with 50 metre and 75 metre buffer zones were studied. The studies were repeated with four different structures of built up areas. First of all, it was noted that the buffer zone cannot be too wide. Inside the zones there is nothing but possibly some corners of the buildings. Also the boundary line in the small resolution seemed to form a serpentine. So the 50 metre resolution with a 100 metre zone was ignored. The second alternative, the 75 metre resolution, was found to be good but the 100 metre resolution proved to be even better after in-depth study. This resolution was just generous enough, the bulk of data decreased, inside cities to a third of the original data, and the outline which took its final form, looked good.

13. A crucial problem was the part of the definition that overthrows the 200 metre distance rule. According to the definition, on the one hand, the distance can exceed 200 metres if the cluster of buildings lies inside the sphere of influence of a major urban settlement and, on the other hand, the distance could be less than 200 metres when the border between localities and the sparsely populated area is unclear. After several tests and many printouts, the solution to this problem was quite simple. It was found that all areas not more than 100 metres from the boundary line were the ones required. Primarily, they were on the side of outgoing roads of a locality. Because of good accessibility and short distance, those areas were considered to be influenced by and in such close interaction with the main area that it was justifiable to merge them into the locality.

III. EVALUATING THE METHOD

14. To evaluate the adopted method, the effect of the chosen method on the whole work process had to be analysed and the results of the method produced had to be assessed.

III.1 Effect on the working process

15. After choosing ArcInfo and the tools inside it, all attention was focused on the functions and commands, their performing order and parameters, instead of verifying algorithms and validity of several programs. Less time and resources were required for developing the process. Through gridding and buffering, the bulk of data decreased considerably (see Table 1). The whole process, delineating, coding and naming, was finalised within two weeks. In conclusion, the method is very efficient in terms of costs and resources.

Type of features	Number of features
Buildings	1 270 000
Grid cells	680 000
Buffered polygons	187 000
Localities	747

Table 1. Quantity of data during the process

III.2 Evaluating the results

16. Because the process is incomplete at this stage, there exist only a few possibilities to evaluate the results, i.e. by verifying areas as a map element for the purpose of visualising and by computing a single indicator to describe how well the method follows the description of localities.

17. The indicator in question is made to measure the distance between buildings inside the boundaries. It seems that inside localities the structure of built-up areas is quite dense and the areas with a strong influence on the main area do not seem to have not so much effect on the indicator (Table 2).

Locality		Buildings	Nearest neighbour (meters)		
ID	Name	Frequency	Mean-distance	std-distance	Max-distance
2088	Pasina	100	40,8	37,8	336,1
0309	Tarvasjoen kt.	300	49,9	43,3	284,4
1011	Paltamon kt.	661	47,6	39,3	323,8
0242	Merikarvia	807	57,5	44,1	386,3
0937	Viitasaarem kt.	1007	40,1	31,9	349,8
0058	Kirkkonummen kt.	2179	42,4	31,1	357,8

Table 2. The density of buildings inside localities

18. As pointed out, the outline the method formed was quite satisfactory. Also, the accuracy of the method seems to be good for the purpose of visualising. Localities are often drawn on the scale of 1:200,000 or smaller. However they can be safely used even on scales 1: 100 000 or 50 000.

19. The coverage of uninhabited or very sparsely populated areas inside localities was formed mainly for the purpose of visualising. This was done by overlaying the coverage of the final localities and the polygons that were generated by buffering the built-up grid cells. Only areas large than one hectar are included in the data. The coverage of "gaps" as well as localities are intersected with a watermask to

calculate land- and water areas. The watermask used is on the scale of 1:50 000 – 1:200 000 and is called Water element, Finland. The owner of that data is National Land Survey of Finland. Also, the number of inhabitants is calculated to the “gaps” so, by overlaying those two coverages, it is possible to cut the “gaps” off the locality and thus gain more accuracy (Figure 5).

20. It is important to notice that when a grid is created, an origin can be given for it. By using the same origin or locating it by calculating multiples of a grid size, it is certain that the grids form areas in exactly the same way and to the same location. So, in order to delineate former or future localities, they must be comparable to each other.

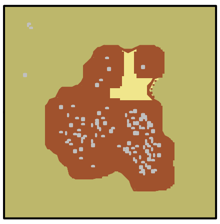

Figure 5. Locality and its unbuilt areas overlay

REFERENCES

ArcDoc 8.0. 1998. Environmental Systems Research Institute, Inc (ESRI). USA.
Tomlin, D. 1990. Geographic Information Systems and Cartographic Modeling. Prentice Hall, Englewood Cliffs, N.J. USA.
Taajamat 1995, Väestölaskenta 1995. Osa 4. Tilastokeskus. Helsinki 1997.